Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

STANDARDIZATION OF ARITHMETICAL WORK.

By W. L. Vosburgh.

For an understanding of the scope of this discussion, it seems necessary to define at the outset the term "standardization" as herein used. "Standardization of arithmetical work" is to be considered from the point of view of modern educational psychology, viz., that its basis be the facts of arithmetical experience rather than mere opinions regarding the same. This view implies the possibility of establishing objective scales for measuring arithmetical abilities.

A word as to the distinction between ***subjective*** and ***objective*** scales of measurements for mental abilities seems necessary and essential at this point. The explanation of this distinction is taken from Thorndike's "Educational Psychology," page 10: "A measurement of human nature to be useful for our purpose must identify the ***amount*** in question for any competent thinker, just as a useful description must ***identify*** the ***object*** in question. To do this it must be objective, that is, free from individual caprice, so that any competent person making the same measurement would get the same result. (Not of course exactly the same. There is a personal equation in even the most objective measures, such as the length of this line ————————. If they measured it to thousandths of a millimeter, competent observers would not get the same result, except by chance. Nor would the same observer in several independent measurements. The ultimate distinction between objective and subjective is simply that in the former sort of measurements competent observers use very nearly the same criteria and, though independent, agree very closely, whereas in the latter they use very different criteria and, if independent, agree only roughly. Objective measures may be defined as measures which competent observers could repeat and verify or reject, and subjective measures as measures which they could not so repeat and verify.)"

The purpose of this paper is to arouse an interest in a scientific study of the teaching of arithmetic, hoping thereby to encourage the teachers of arithmetic to measure the results of their teaching objectively—a more definite and scientific way than is possible by means of the ordinary form of test or examination, and to secure their co-operation in the work of establishing objective scales of measurement for all arithmetical abilities.

DESIRABILITY OF OBJECTIVE STANDARDS.

The desirability of objective standards in all school work, and particularly in arithmetical work, is I think obvious to all of us; the problem with which we are concerned is the possibility of establishing such standards in arithmetical work. Before considering the latter however I will quote from two authorities as to the former.

Dr. Hanus in his paper on "Our Chaotic Education" which he read before the Department of Superintendence of the N. E. A. at its meeting in February, 1902, said as follows: "Now, just as we have not organized and adequately assimilated a generally accepted educational doctrine, so we are without a body of recorded educational experience. Results actually achieved and collectively presented constitute a force that is capable of sweeping away superficial criticism or paralyzing scepticism on the one hand, and meddlesome interference and impatient clamoring for premature results on the other. Isolated successes have been advertised, to be sure, and failures, more or less obvious, have sometimes been frankly confessed, and sometimes unwisely suppressed. But in neither case have we had an orderly presentation of both successes and failures over a wide area. We have had plenty of experiments; indeed, as I have intimated, our whole educational activity for nearly a generation has consisted of experiments. Just as every educational theorizer has worked by himself without taking due account of the labors of his fellow-workers in the same field, so every superintendent has pursued his way, apparently in blissful indifference to what his fellow-superintendents were doing, multiplying instances and varying conditions *ad libitum*. How is it possible to extract any confirmation of alleged results from such a heterogeneous procedure? And we never can get such confirmation until we abandon our absurd extreme of individualism in these experiments and work together for the attainment of the same ends. . . .

" Briefly, then, we must organize our educational experience just as we must organize our educational doctrine, if we are to make real progress."

He then gives two or three illustrations of what he means; the one of interest to us is as follows: " Suppose the conditions to be something like this: Five hundred or a thousand pupils in each city to begin the study of arithmetic in the first year, and a similar number to begin it in the second year, and a third similar group to begin it in the third year of school. At the end of the sixth year of school compare the attainments of the three groups of pupils. Would not the conclusions reached by such an experiment have a convincing value which no amount of assertion beginning 'in my schools,' or 'so far as my experience goes,' or 'I believe,' or 'in my opinion' could possibly have? . . . I do not believe that such co-operation is impossible. Why should it be? Experiments similar to those suggested are everywhere in progress; co-operation in large enterprises of all kinds is possible. Why should it be impossible in education only?

"Under such circumstances we could face the teaching profession and the general public with facts, instead of opinions. The enormous difference between the weight of these two very different things in educational affairs still remains to be experienced. . . ."

The second authority on the desirability of objective standards in school work, Dr. E. L. Thorndike, the inspiring genius of the quantitative study of education in American today, said in 1904: "The study of education is beginning to be quantitative. We are no longer satisfied with vague arguments about what this or that system of administration or method of teaching does, but demand exact measures of the achievement of any system or method or person. We are becoming properly disgusted with the one-sided book-keeping which only takes account of the dollars spent, and neglects the debit side, the income in knowledge, habits, power, zeal, and ideals.

" This ambition toward an exact objective measurement of the results of educational endeavor is a symptom of healthy scientific fervor and also of common-sense wisdom. It is in every respect and for every reason right that the same painstaking precision should be applied to the study of human welfare that is now given to the study of Jupiter's satellites or to the petals of the daisy. It is right, too, that to a general faith in the efficacy of systematic education there should be added a cold-blooded inquiry into just what comes of it all. No one possessed of either science or sense will deny the value of successful quantitative study of school work.

"The complexity of the phenomena of education is obvious and need not be discussed here. The lack of commensurate units in which to measure intellectual and moral qualities has been by far the greatest barrier to quantitative science, all the more so in education because we have been unconscious of the lack. The units at hand are defective in that they are subjective, depend on individual caprice instead of universal agreement, are not commensurate, and are reckoned from unknown zero-points. In even so simple a matter as ability in arithmetic, the measure to be given to any individual is not an easy affair. In the first place, the marks will vary somewhat with the individual who examines the examples done; secondly, the credit to be given for any example as compared with any other is unknown; and, next, even if the test comprised twenty examples known to be of equal difficulty and to have been marked impartially, the fact that one boy scored 14 and another 7 need not mean that the former is twice as far above zero in arithmetical ability as the latter."

Possibilities of Objective Standards.

Our problem, the possibility of standardizing arithmetical work, is summed up by Dr. Thorndike in the following: "To have in education the real benefits of quantitative science, we must spend arduous years in devising, testing, and standardizing units of measurement, in searching for convenient arbitrary zero-points, and eventually for real zero-points, and in determining the sources and amounts of the errors of measurement. It has taken many centuries and the labors of many gifted men to develop the present system of physical measurements, and the task is far harder in the case of intellectual and moral facts.

That the adequate measurement of intellectual and moral facts is difficult does not, however, at all imply that it is impossible or that inadequate and erroneous measurements are not useful. . . . Inadequacy and error are relative terms. It is a great advance to regard the sun as ten times farther away from the earth than the moon is, if the previous belief was that it was only half as far. . . . Only the ignoramus regards any measurement as perfect. All are approximations, and a very rough approximation may be far more accurate than previous measurements."

We are fortunate in having, at the present time, three very definite and specific contributions to the scientific study of arithmetical abilities. These are the studies by Rice ('02), Stone ('07), and, Courtis ('10 and '11), respectively.

Dr. J. M. Rice, formerly editor of *The Forum,* may be regarded as the man in this country to take the initiative in an attempt to study scientifically the actual achievements of the schools in the school arts (his studies investigating the teaching of spelling, arithmetic, and language), with a view to establishing just how much time can be saved by a suitable restriction and selection of subject matter.

Dr. Rice in 1902 measured the arithmetical abilities of some 6,000 children of grades 4, 5, 6, 7, and 8, in eighteen different schools, in seven different cities. His tests for each of these grades and a discussion of the results attained by the various systems may be found in the *Forum,* Vol. 34.

We are concerned here primarily with: (1) the abilities tested, (2) the variation in the results attained within the systems and between the systems, and (3) the actual contribution of his study to the establishment of objective standards in arithmetical work.

The test itself consisted of eight examples for each grade. The examinations were given to the pupils of the fourth, fifth, sixth, seventh, and eighth grades. They were not given below the fourth grade because he aimed to discover what children are able to do on leaving school. For the purpose of studying the growth of mental power from year to year, some of the problems were carried through several grades. Thus, of the eight questions for the fourth grade, five were repeated in the fifth, and three in the sixth, etc. Moreover, this repetition enabled him to see not only how the results in the fifth and sixth grades, in regard to certain problems, compared with those of the fourth in the same school, but also how the results in the fourth grade of some schools compared in these examples with those of the fifth and sixth grades of others, etc. The subject matter of each test was such as to attempt measure, at the same time, ability in both reasoning and fundamentals.

The range of variations in the results obtained on these tests was very wide, viz., in the seventh year the class averages range from 8.9 per cent. to 81.1, and in the eighth year from 11.3 to 91.7. The averages for schools taken as a whole varied between 25 and 80 per cent.

The general average for all schools examined was, in round numbers, 55 per cent., made up as follows: 60 per cent. for the fourth year, 70 per cent. for the fifth, 60 per cent. for the sixth, 40 per cent. for the seventh, and 50 per cent. for the eighth. In view of what the satisfactory schools have shown, Dr. Rice thinks that 60 per cent. for the fourth grade, 70 per cent. for the fifth, and 60 per cent. for the sixth are reasonable expectations. However, 40 per cent. for the seventh grade and 50 per cent. for the eighth are too low, as these figures are not at all representative of what the successful schools have been able to accomplish, but result from the fact that, in the majority of instances, the seventh and eighth grades were lamentably weak. As the average for the seventh grade of the five successful schools was 61.4, and that of the eighth grade 77.2, he thinks that less than 50 per cent. for the seventh year and 60 per cent. for the eighth should not be regarded as satisfactory.

Dr. Rice says: "In suggesting a standard, it is, of course, understood that the figures mentioned in the last paragraph would only be applicable to an examination whose degree of difficulty were the same as my own. Teachers desirous of knowing how their pupils compare with those of other schools could try the questions as I have given them, or, if they feared that the publication of the problems had diminished the value of the test, they might change the figures without altering the degree of difficulty. However, in due course of time there ought to be no difficulty in establishing standards in arithmetic with mathematical precision. This may be quite readily done by selecting types of examples and determining by research what percentage ought to be obtained on each of them by the class for which they are intended. When this point has been reached, a standard will also have been fixed for a combination of examples of various degrees of difficulty."

In addition he gives us the following conclusions in regard to twelve possible factors of school work. These are based upon the facts revealed in the results attained by the different school systems. His conclusions are briefly summarized, by Stone, as follows:

(1) *Home environment:* "As in spelling so in arithmetic, this mountain, on close inspection, dwindles down to the size of a molehill"; (2) *Size of classes:* "No allowance whatever is to be made for the size of the class in judging the results of my test"; (3) *Age of pupils:* "This factor can be held accountable for the difference shown to only a slight degree if at all"; (4) *Time of day:* "This likewise is not a determining factor"; (5) *The time devoted to arithmetic in the school:* "A glance at the figures will tell us at once that there is no direct relation between time and result"; (6) *The amount of home work required:* "This fact is also denied to be a controlling one, as the highest five schools had practically abandoned home work, while the lowest ranking city required most"; (7) *Method of teaching:* In the schools that passed the tests satisfactorily no special method had been in use; (8) *Teaching ability:* "Teachers of most successful schools had no better ability than those of unsuccessful schools"; (9) *Course of study:* By a curious line of argument Dr. Rice reaches the conclusion that the course of study is not a factor to be considered because the tests were fair to pupils who had been taught by all courses of study; (10) *Superintendent's training of teachers:* This was as much in vogue in localities that did poorly as in those that did well; (11) and (12) *Superintendents' establishing of standards and testing for results:* These Dr. Rice concludes to be the large and controlling factors.

In conclusion Dr. Rice says: "By reason of the high percentages obtained in certain schools, laboring under ordinary conditions, we must accept as a fact that nearly all children can be trained to solve any ordinary problem in arithmetic, based upon principles they have studied. Consequently, if the normal child is not reasonably proficient in that branch, as far as he has advanced in it, the fault is not his."

The second contribution to the quantitative study of arithmetical work is that of Dr. C. W. Stone made in 1907. The results of this study were published in 1908 as Vol. 19 of Columbia University Contributions to Education (Teachers College Series) entitled "Arithmetical Abilities and Some Factors Determining Them." This study took account of, and profited by, that made by Dr. Rice. Its improvements over the former are in the data used, and in more refined and scientific methods of gathering and handling the same.

"The central purpose of this study" (to quote the author), "broadly stated, is to make one more contribution to the exact knowledge of the relation between distinctive educational procedures and the resulting products."

The *method* of this inquiry may be briefly characterized as the application of the statistical method to mental measurements. The three principal measurements made are the arithmetical abilities of 6A pupils, the time expended, and the course of study materials used in securing these abilities.

The sources of data were 26 school systems in six different states from Indiana to Massachusetts. Dr. Stone visited and personally gave, under conditions as nearly identical as possible, two tests, one in fundamentals and one in reasoning, to 3,000 6A grade children. The time for the test in fundamentals was set at 12 minutes; for reasoning at 15 minutes, each test being purposely too long to be finished by any pupil within the time limit (14 examples in fundamentals, 12 in reasoning).

The main purpose of the test in fundamentals was to determine ability of 6A grade pupils in addition, subtraction, multiplication and division. The main purpose of the test in reasoning was to determine the ability of 6A grade children to reason in arithmetic. The test in fundamentals was so arranged as to enable the pupils to meet the main difficulties in the first six problems; in both tests the problems were arranged in order of difficulty with the less difficult first.

The variations in abilities between the systems in the grade tested (6A) were as follows: In reasoning the points scored varied from 356 to 914, with an average deviation of 112, the median being 551; in fundamentals from 1,841 to 4,099, with an average deviation of 421, the median being 3,111. Translated this means that if we take the efficiency of the school making highest score in reasoning as 100 per cent. the median represents an efficiency of 60 per cent., the lowest of 39 per cent.; in fundamentals the median score represents an efficiency of 76 per cent., while the lowest 45 per cent. The variability among individuals within a system was even greater than that among systems. Among 500 pupils chosen at random from four representative public school systems: In fundamentals they varied from 0 to 15.2: 9 out of 500 failed to solve any example in the reasoning test: The median score was about 6: 33 made score of 2 or less. In one system 15 out of 100 pupils made no score in subtraction and 29 made one each. In general, pupils varied most in division, next in reasoning and about equally in addition and multiplication. In regard to the fundamentals, from their coefficients of correlation, he tentatively states the following: "that the possession of ability in addition is the least guarantee of possession of ability in others; that the possession of ability in multiplication is the best guarantee of the possession in others; and that this probably means that multiplication is like addition on its mechanical side and like division on its thinking side. Hence if one wished to measure abilities in fundamentals by a single test, one in multiplication would be best; and a test in division would probably be the best single measure of arithmetical abilities." As to the lack of correspondence between serial standings in time-cost and serial standing in abilities of systems we find the system devoting the smallest per cent. of time to arithemtic (7 per cent.) was third in average standing of comparative achievement; the system devoting largest per cent. of time to arithmetic (22 per cent.) was twelfth; the one devoting 14 per cent. was lowest of all.

CONTRIBUTIONS.

His conclusions may be summarized as follows: (1) That the general method of this research is a means by which hypotheses are to be tested and opinions to become facts. (2) That *diversity* sums up the findings of this study, as shown by: (I.) A variability of scores among systems of 356 to 914 points, with an average deviation of 112 in reasoning; and 1,841 to 4,099, with an average deviation of 421 in fundamentals; (II.) A variability of mistakes among systems of 14.5 per cent.

to 4.7 per cent. of all the steps attempted in addition, and 45.1 per cent. to 14.4 per cent. of all the problems attempted in reasoning; (III.) A variability of 507 to 1,854 week-minutes with an average deviation of 222 week-minutes in time expenditure, without home study; and 507 to 2,179 week-minutes with an average deviation of 269 week-minutes, including home study; (IV.) A variability in average ratio of time to abilities of from 2.26 to .64; (V.) A difference in course of study excellence which can hardly be put in words. (3) The greatest need shown by the research is standards of achievement. That the great variability herein shown would exist if school authorities possessed adequate means of measuring products is inconceivable; and it is believed that the present study will help standardize the work in arithmetic for the first six grades. (4) Finally that there is no *one* factor that produces abilities, there is no single *summum bonum* in teaching arithmetic.

The third contribution to the quantitative study of arithmetical work is that of S. A. Courtis, an account of which may be found in Vols. 10 and 11 of *Elementary School Teacher;* the latest report of the progress of his study may be found in the *Elementary School Teacher* for November, 1911. In this latest study he has secured the co-operation of teachers in England and Germany as well as in the United States. Mr. Courtis began his study of the measurement of growth and efficiency in arithmetic, in 1908, in the school with which he is connected (The Liggett Home and Day School of Detroit, a private school for girls). His ambition was to find the place of his school among the 26 school systems reported by Dr. Stone. The results secured by his school on the Stone tests and an analysis of the same are given in the *Elementary School Teacher,* Vol. 10. In speaking of the results the test yielded he says: "It has shown the need for greater knowledge of what is actually taking place in the child's mind as he passes through grade after grade. It has made plain the uncertainty of the product of the present educational system as well as the value of a complete and many-sided training in the early grades if there is to be great ability in the later grades. More than all else it has proved, conclusively to the writer at least, that it is practicable to measure, not only the general condition of arithmetic teaching throughout a school, the growth in ability and efficiency from grade to grade, the defects and needs of any one grade or individual, but the *effects of changes in method or procedure as well.* By a series of tests, through a number of years, it ought to be possible to build up a real science of teaching and to determine by strict experimental methods the truth or falsity of any educational hypothesis." With this idea in mind Mr. Courtis devised, the following year, 1909, eight tests designed to measure the change produced by a year's regular work under existing conditions in order to have some standard by which to measure the effect of future changes in method. Because of the complexity of the problem, the scope of the investigation was limited to the simplest work with whole numbers in the four fundamental operations. The specific abilities to be measured by the tests were: (1) ability to give a ready response in the case of any of the elementary combinations in the four operations, (2) ability to recognize a situation in a problem as calling for the use of a certain operation, (3) ability to borrow and carry, and (4) the ability to copy correctly the figures of a given example or problem. Accordingly provision was made for the testing of each of these as well as the general abilities involved in working abstract examples and concrete problems.

The subject matter of the tests, eight in number, was as follows:

Test No. 1, Addition,
Test No. 3, Multiplication,
Test No. 2, Subtraction,
Test No. 4, Division,
(Combinations, 0–9),
Test No. 5, Copying figures (rate of motor activity),
Test No. 6, Speed reasoning (simple one-step problems),
Test No. 7, Fundamentals (abstract examples in the four operations),
Test No. 8, Reasoning (two-step problems).

The nature of the investigation made it essential that the tests should be given to all grades from the lowest to the highest under identical conditions.

The quantitative studies of the teaching of arithmetic in his school during years 1908–09, 1909–10, showed Mr. Courtis the possibilities of measuring accurately the changes produced by the teaching effort. Clearly defined standards alone were lacking. His latest study therefore deals with an attempt to determine "standard scores in arithmetic" and may be found in *Elementary School Teacher,* November, 1911.

In this study he has already secured the co-operation of some 60 or 70 different schools, in 10 different states, from Kansas to New Hampshire. More than 18,000 sets of his tests have been distributed by him and the co-operation of teachers in England and Germany in this study is already assured. Last June more than 9,000 individual scores, from 60 to 70 schools, were returned for tabulation.

The subject matter of these tests has been considered; as to the variations in abilities shown among the systems we have: Of 1,401 eighth-grade children measured as to their knowledge of the multiplication tables, 2 could write answers at a rate of but 5 answers a minute, 17 at the rate of 15 answers a minute, 180 at 25 per minute, and so on, the average for the entire group being 43.

It is particularly to be noted that the wide range of the distribution is caused not so much by differences between schools as by great variation in the abilities of individual children. The amount of variation in the grade averages of different schools does not equal the variation in the scores of the members of any one class.

The fact shown is a constant feature of the returns for all the tests and grades, and can only mean that the differences in the abilities of individual children are greater factors in determining relative rank in school work than all the differences in abilities of teachers, courses of study, or methods of work combined. The result is that the grades in all the schools overlap to an extent well nigh incredible.

Courtis's contribution to the standardizing of arithmetic is significant in that he suggests standard scores for the abilities tested, for each grade from III. to IX. inclusive. Such standard scores may be the average scores obtained on each test by all pupils tested in each grade. These are given in Table III. of the November article, p. 133; or a standard score may be derived as follows: "For each test an eighth-grade score was selected such that it was equalled or exceeded by 30 per cent. of the eighth-grade children measured. This score was plotted and a smooth curve drawn, having the same general form as the average curve and coinciding with it in the lower grades. The scores for each of the other grades were then determined from the graph." Standard scores derived on this basis are given in Table IV., p. 135. Either of these standard scores may be used by teachers as definite objective standards toward which to work, *e. g.*, "at the end of a year's work, an eighth-grade child should be able to copy figures in pencil on paper at the rate of 117 per minute; to write answers to the addition combinations at rate of 63 per minute, subtraction, multiplication and division 49 per minute, etc., etc. At the present time 70 per cent. of the eighth grade cannot meet these standards but 3 per cent. of the fifth grade can.

A review of each of these three studies has been made with the hope of enlisting the sympathies of those assembled here in this movement tending toward the standardization of arithmetical work, and of making it apparent that as a foundation for such work we already have accumulated a well-defined and important body of facts of arithmetical experience. The value of these to all future quantitative studies of arithmetical abilities cannot be overestimated.

As to the part that each one of us may take in this movement, we may consider any one or more of the following suggestive lines of work:

1. That the tests and examinations now given be used in a scientific way and be made to be careful scientific measures of changes in our pupils. Dr. Thorndike says that "if an examination, instead of being a hasty, subjective selection of questions graded still more personally (and alas, how hastily), were made a serious educational measurement, the examination papers of a year would give us a large start toward knowledge of what (science) teaching actually does."

2. That we make use of the tests devised by Rice, Stone, or Courtis, as objective measures of achievement of our pupils and objective measures of the results of the teaching effort.

3. That there is an opportunity for co-operation with Mr. Courtis at the present time in his study, recently undertaken, to determine accurately the proper amount and form of drill re-

quired to produce the standard abilities suggested, and similar determinations of standard yearly growths.

4. That other and similar tests be devised for measuring objectively specific abilities in arithmetic and the results of our teaching.

In conclusion, a sample test, designed by myself to measure the ability of entering Normal students in addition will be submitted for your inspection.

BROCKPORT, N. Y.